Complex histories of gene flow and a mitochondrial capture event in a nonsister pair of birds

Andersen, Michael J.; McCullough, Jenna M.; Gyllenhaal, Ethan F.; Mapel, Xena M.; Haryoko, Tri; Jonsson, Knud A.; Joseph, Leo

Published in: Molecular Ecology

DOI: 10.1111/mec.15856

Publication date: 2021

Document version
Publisher's PDF, also known as Version of record

Document license: CC BY-NC-ND

Citation for published version (APA):
Complex histories of gene flow and a mitochondrial capture event in a nonsister pair of birds

Michael J. Andersen¹ | Jenna M. McCullough¹ | Ethan F. Gyllenhaal¹ | Xena M. Mapel¹,² | Tri Haryoko³ | Knud A. Jønsson⁴ | Leo Joseph⁵

¹Department of Biology and Museum of Southwestern Biology, University of New Mexico, Albuquerque, New Mexico, USA
²Animal Genomics, ETH Zürich, Lindau, Switzerland
³Museum Zoologicum Bogoriense, Research Centre for Biology, Indonesian Institute of Sciences (LIPI), Cibinong, Indonesia
⁴Natural History Museum of Denmark, University of Copenhagen, Copenhagen Ø, Denmark
⁵Australian National Wildlife Collection, CSIRO National Research Collections, Canberra, Australian Capital Territory, Australia

Correspondence
Michael J. Andersen, Department of Biology and Museum of Southwestern Biology, University of New Mexico, Albuquerque, NM, USA. Email: mjandersen@unm.edu

Abstract
Hybridization, introgression, and reciprocal gene flow during speciation, specifically the generation of mitonuclear discordance, are increasingly observed as parts of the speciation process. Genomic approaches provide insight into where, when, and how adaptation operates during and after speciation and can measure historical and modern introgression. Whether adaptive or neutral in origin, hybridization can cause mitonuclear discordance by placing the mitochondrial genome of one species (or population) in the nuclear background of another species. The latter, introgressed species may eventually have its own mtDNA replaced or "captured" by other species across its entire geographical range. Intermediate stages in the capture process should be observable. Two nonsister species of Australasian monarch-flycatchers, Spectacled Monarch (Symposiachrus trivirgatus) mostly of Australia and Indonesia and Spot-winged Monarch (S. guttula) of New Guinea, present an opportunity to observe this process. We analysed thousands of single nucleotide polymorphisms (SNPs) derived from ultraconserved elements of all subspecies of both species. Mitochondrial DNA sequences of Australian populations of S. trivirgatus form two paraphyletic clades, one being sister to and presumably introgressed by S. guttula despite little nuclear signal of introgression. Population genetic analyses (e.g., tests for modern and historical gene flow and selection) support at least one historical gene flow event between S. guttula and Australian S. trivirgatus. We also uncovered introgression from the Maluku Islands subspecies of S. trivirgatus into an island population of S. guttula, resulting in apparent nuclear paraphyly. We find that neutral demographic processes, not adaptive introgression, are the most likely cause of these complex population histories. We suggest that a Pleistocene extinction of S. guttula from mainland Australia resulted from range expansion by S. trivirgatus.

Keywords
introgression, Maluku Islands, mitonuclear discordance, Monarchidae, Symposiachrus, ultraconserved elements
1 | INTRODUCTION

Two inter-related phenomena are increasingly recognized as elements of the speciation process: hybridization and mitonuclear discordance. Hybridization is now seen as a pervasive (Nosil, 2008), almost inevitable, element of speciation (reviews in Edwards et al., 2016; Joseph, 2018; Ottenburgs et al., 2017; Rheindt & Edwards, 2011; Smadja & Butlin, 2011; Taylor & Larson, 2019) and mitonuclear discordance is being increasingly detected in speciation studies (Bonnet et al., 2017; Good et al., 2008; Kears et al., 2014; Pons et al., 2014; Toews & Brelsford, 2012). Consequently, genomic approaches to study both phenomena are now commonplace (e.g., McElroy et al., 2020; Payseur & Rieseberg, 2016). One outcome of hybridization during speciation is introgression (i.e., one-way gene flow through repeated backcrossing). Introgression, in turn, may generate mitonuclear discordance. Bonnet et al. (2017) introduced the term “massively discordant mitochondrial introgression (MDMI)” to describe extreme cases of mitonuclear discordance, whereby all or nearly all individuals of a species have the mitochondrial DNA (mtDNA) of another species but show very little nuclear introgression. This phenomenon is also termed mitonuclear capture. Hybridization and MDMI can be thought of as leading to one genome replacing another (Mallet, 2005).

Massively discordant mitochondrial introgression can arise through multiple mechanisms. First is through a suite of mitonuclearly neutral or nonadaptive processes. These include sex-related asymmetry in interspecific gene flow or negative selection on multiple nuclear loci (discussed extensively in Bonnet et al., 2017), or a geographical range expansion following introgressive hybridization (Currat et al., 2008). Second are adaptive processes that center on positive selection for the introgressed mitochondrial genome. For example, enhanced cellular-level physiology can arise from introgression that places a novel mitochondrial genome in the nuclear genomic background of the introgressed species (Hill, 2019). For theoretical and empirical discussions of these mechanisms, see Borge et al. (2005), Currat et al. (2008), Petit and Excoffier (2009), Rheindt and Edwards (2011), Toews and Brelsford (2012), Toews et al. (2014); Sloan et al. (2017); Hill (2019), and Zhang et al. (2019), Bates-Muller-Dobzhansky incompatibilities represent yet a third mechanism that can lead to mitonuclear discordance. These incompatibilities arise from hybridization between two populations each having coadapted mitochondrial and nuclear genomes. This mechanism, however, is probably less likely to lead to mitochondrial capture (see Burton & Barreto, 2012; Burton et al., 2013). The introgressing and introgressed species need not be sister species and mitochondrial capture can occur long after speciation has been completed (Drovettski et al., 2015; Kears et al., 2014; Shipham et al., 2016). In addition, two other factors could explain mitochondrial discordance: incorrect taxonomy and occasional gene flow, either historical or contemporary. Here, we explore these hypotheses and their difficulties in a case of mitochondrial discordance in birds.

An opportunity to study massively discordant mitochondrial introgression is offered by two nonsister species of Australasian monarch-flycatchers. The Spectacled Monarch Symposiachrus trivirgatus (Aves: Passeriformes: Monarchidae) comprises eight geographically disjunct subspecies in three regions (Figure 1; Beehler & Pratt, 2016; del Hoyo & Collar, 2016; Schodde & Mason, 1999): (a) in the Lesser Sundas and the Maluku Islands of Indonesia, (b) in the Louisiade Archipelago, south-east of mainland New Guinea, and (c) in northeastern Australia, Torres Strait Islands, and southern New Guinea. Within Australia, three subspecies are currently recognized (see Beehler & Pratt, 2016; and Schodde & Mason, 1999 for details; Figure 1). Symposiachrus t. albiventris is sedentary and occurs in northernmost Cape York Peninsula and the Torres Strait Islands. Sedentary S. t. melanorrhous and migratory S. t. gouldii have largely allopatric breeding distributions in north-eastern Australia but overlap in the nonbreeding season. A nonsister species, the Spot-winged Monarch S. guttula, ranges across all of New Guinea and some of its eastern and western satellite islands. Symposiachrus guttula does not occur in Australia (Andersen et al., 2015); the Torres Strait narrowly separates the two species. They co-occur in a small part of southern New Guinea when migrant S. trivirgatus from Australia are present seasonally. They closey approach each other, however, on outlying islands east and west of New Guinea. Phenotypically, S. trivirgatus males from Indonesian populations resemble S. guttula.

Earlier studies that included up to six individuals of S. trivirgatus identified two nonsister haplogroups in that species’ mtDNA (Andersen et al., 2015; Filardi & Smith, 2005). One haplogroup was sister to mtDNA of S. guttula and the other was phylogenetically separated from it by several other Symposiachrus species (Andersen et al., 2015). Prior to the present study, more extensive sampling of Australian S. trivirgatus individuals (reported below) expanded the numbers of S. trivirgatus individuals in these two haplogroups and affirmed their phylogenetic relationships and composition being solely of S. trivirgatus. The key point here is that neither haplogroup of S. trivirgatus mtDNA included any individuals of S. guttula. This is potentially consistent with the “allophyly” stage of incomplete lineage sorting (Omland et al., 2006); however, the nonsister species relationships of the two species coupled with one S. trivirgatus haplogroup being sister to one lineage of S. guttula rule out this explanation.

The aims of this study were to further investigate the apparent massively discordant mitochondrial introgression of S. trivirgatus with additional sampling of all S. trivirgatus subspecies and S. guttula. We tested four hypotheses that might explain apparent MDMI in Australian S. trivirgatus: cryptic speciation, gene flow, neutral capture, and adaptive capture. We applied phylogenetic, population genetics, and demographic modelling approaches of both nuclear DNA and mitochondrial genomes. Cryptic speciation is considered because Indonesian and Australian S. trivirgatus may not be closely related (del Hoyo & Collar, 2016; Eaton et al., 2016). It predicts that species-specific clades of nuclear DNA diversity will mirror the pattern of relationships among mtDNA haplotypes. We applied tests for signals of contemporary and historical gene flow across all of our samples and used single nucleotide polymorphisms (SNPs) derived from ultraconserved elements (UCEs) in the nuclear genome and
whole mitogenomes of samples spanning the ranges of both species. Relationships among component populations and subspecies of *S. trivirgatus* are also reported. Ultimately, we hope to increase our understanding of speciation in the biologically important Indo-Pacific region.

2 | MATERIALS AND METHODS

2.1 | Sampling and DNA extraction

Our data set comprised 81 ingroup taxa (57 *S. trivirgatus sensu lato* and 24 *S. guttula*; Table S1). We sampled broadly across the distributions of *S. guttula* and *S. trivirgatus* (Table S1, Figure 1), including all eight subspecies of *S. trivirgatus* (Clements et al., 2019; del Hoyo & Collar, 2016; Dickinson & Christidis, 2014). All mainland New Guinea samples were from Papua New Guinea because we lacked tissue samples from the west, but for simplicity, we refer to this population as “Mainland New Guinea.” We rooted phylogenetic analyses to *Myiagra ruficollis* (sequenced for this study), but used a more closely related outgroup, *Symposiachrus verticalis*—from Moyle et al. (2016)—for analyses of gene flow. Most samples were derived from frozen or ethanol-preserved, specimen-vouchedered tissues; however, three samples of Indonesian island populations of *S. trivirgatus* were sourced from specimen toepad clippings (Table S1). For tissue-derived libraries, we extracted genomic DNA following standard protocols with a Qiagen DNeasy Blood and Tissue Kit and quantified extracts with a Qubit 3.0 Fluorometer (ThermoFisher Scientific).
2.2 Target capture of ultraconserved elements

We performed target capture of UCEs (Faircloth et al., 2012) at the University of New Mexico. Prior to library preparation, we visualized DNA extraction quality via gel electrophoresis to determine shearing protocol. We sheared 250 ng of genomic DNA to 500 base pair (bp) fragments with a Covaris M220 focused-ultrasonicator (50 W peak incident power, 10% duty factor, and 200 cycles per burst for 55–75 s, depending on extraction quality). We prepared Illumina libraries using 1/2 the volume per reaction of the Kapa Biosystems Hyper Prep kit and the iTruStub dual-indexing system (baddna.org), following established protocols (Faircloth et al., 2012). We enriched pooled libraries (500 ng of 8 equimolar samples per pool) for 24 h at 65°C using the Arbor Biosciences (formerly MYcroarray) myBaits tetrapod 5 K v1 probe kit, targeting 5060 UCE loci. UCE-enriched pooled libraries were quality checked with a bioanalyzer (Agilent Bioanalyzer 2100), qPCRRed, and combined at equimolar ratios prior to sequencing on a single lane of an Illumina HiSeq 3000 with a PE-150 flow cell at the Oklahoma Medical Research Foundation (OMRF). We handled the six toepad-derived samples differently than the tissue-derived samples. This is because DNA from museum study skins tend to be fragmented (McCormack et al., 2016; Mundy et al., 1997; Tsi et al., 2019). For toepads, we followed protocols we have used previously (Andersen et al., 2019; McCullough et al., 2019). Toepad-derived libraries were pooled with samples from other projects and sequenced at OMRF using the same sequencing parameters as the tissue-derived libraries.

2.3 UCE data processing and phylogenetic analysis

We processed UCE data, built contigs, trimmed, and aligned loci with the Python bioinformatics package “phyluce” v1.5.0 (Faircloth, 2016; see https://github.com/faircloth-lab/phyluce). First, we demultiplexed raw reads from BCL files using BCL2FASTQ v2.17.1.14 (Illumina Inc.) then removed adaptor sequences and low-quality bases with illumiprocessor v1 (Bolger et al., 2014; Faircloth, 2013; Lohse et al., 2012). We used TRINITY v2.0.6 (Grabherr et al., 2011) to assemble contigs from cleaned reads then extracted contigs for each taxon that matched UCE loci. We aligned each locus specific data sets with MAFFT (Katoh & Standley, 2013) and trimmed them with GBLOCKS v0.91 (Castresana, 2000). We used default parameters for gblocks except for the minimum number of sequences allowed for flank positions, b2, which we set to 65%. These steps produced a final 75%-complete UCE data set in which all locus-specific alignments had data from at least 61 of the 81 individuals.

Prior to phylogenomic analysis of UCE data, we selected the optimal partitioning schemes and substitution models with PARTITIONFINDER2 v2.1.1 using the rcluster algorithm (Lanfear et al., 2014, 2017). Then, we analysed the partitioned, concatenated matrix with RAxML v8.2.4 (Stamatakis, 2014) using the CiPRES Science Gateway (Miller et al., 2010). We evaluated support of the ML topology with 900 rapid bootstrap replicates. To take into account gene tree heterogeneity in our UCE data set, we inferred species trees using SVDQuartets (Chifman & Kubatko, 2014) invoked in PAUP v4.0a (build 163; Swoford, 2003). For this analysis, we considered the nine clades that we inferred from mitochondrial data (i.e., the nine coloured clades in Figure 1b; see mitogenome section below), we evaluated 139,492 quartets, and assessed support with 1,000 bootstraps.

2.4 Mitogenome data processing and analysis

We recovered off-target sequencing reads of the mitochondrial and generated mitogenomes by mapping the illumiprocessor cleaned reads to a mitogenome reference of *Terpsiphone atrocaudata* (GenBank: KT901458) in Geneious R9.1 (Biomatters Ltd). We checked these assemblies against known mitochondrial sequences that were generated with standard PCR and Sanger sequencing (Andersen et al., 2015; Filardi & Moyle, 2005; Filardi & Smith, 2005) to guard against the presence of NUMTs. We also checked these alignments by eye for internal stop codons and failing to find any, we continued with our analyses. We inferred an unrooted haplotype network of the ND2 gene using the unrooted TCS (Clement et al., 2000) algorithm with default settings in PopART (Leigh & Bryant, 2015). We analysed the 13 mitochondrial protein-coding genes using three partitions, one for each codon position. To expand our mtDNA alignment, we downloaded from GenBank 28 additional ND2 sequences of *Symposiachrus* spp., one *Monarcha castaniaventris*, and one *T. atrocaudata*. This additional sampling included all samples from Andersen et al. (2015), plus *S. menckei* (GenBank: KP726921) and *S. trivirgatus bimaculatus* (GenBank: KP726923). We used this expanded mitogenome matrix of 113 samples to estimate a phylogeny in BEAST v2.5.2 (Bouckaert et al., 2014). We assigned the HKY+I + G model to each partition and used a strict clock model and a birth-death tree model. We ran four independent MCMC chains of 10 million generations each, sampling every 2,000 generations. We assessed convergence of parameter estimates and determined that ESS values were all >200 by visualizing trace files in TRACER v1.7.1 (Rambaut et al., 2018). We discarded the first quartile of MCMC generations before summarizing the remaining trees in a maximum clade-credibility tree using LOGCOMBINER v2.5.2 and TREEANNOTATOR v2.5.2 (Bouckaert et al., 2014).

2.5 SNP calling from UCEs

We called SNPs from UCEs to determine individual population assignment and explore gene flow between *S. trivirgatus* and *S. guttula*. We largely followed the seqcap_pop pipeline (Harvey et al., 2016). We de novo assembled our cleaned reads from 15 individuals into a consensus reference assembly using VelvetOptimiser (Zerbino & Birney, 2008). We selected these specific individuals for their geographic and taxonomic breadth within our data set and because they represented samples with the highest number of cleaned reads (see
Table S1). To avoid including markers of differing ploidy, we used BLASTN v2.2.31 (Zhang et al., 2000) to identify Z-linked loci from the Zebra Finch (Taeniopygia guttata) reference genome from ENSEMBL 95 (Zerbinio et al., 2018) and removed them from our reference assembly using a custom python script. We then mapped contigs to targeted UCE loci, and mapped reads to those contigs using BWA v0.7.7 (Li & Durbin, 2009). To convert SAM to BAM files and sort, clean, add read groups to, call PCR duplicates in, and merge BAM files across individuals, we used PICARD v2.18.17 (http://broadinstitute.github.io/picard/). We also used this program to create a dictionary with the reference assembly and SANTOOLS v1.5 (Li et al., 2009) to index the BAM files and reference. With GATK v3.8.1 (McKenna et al., 2010), we called indels using RealignerTargetCreator and realigned them using IndelRealigner. Subsequently, we called SNPs with UnifiedGenotyper, annotated variants with VariantAnnotator, removed variants with a quality score of less than 30 or a quality by depth <5 with VariantFiltration, and finally phased the data with ReadBackedPhasing to produce a VCF file for downstream analyses. With VCFTOOLS v0.1.15 (Danecek et al., 2011), we restricted the data set to biallelic SNPs with no missing data and used custom python scripts to remove possible paralogs based on a cutoff of ≥0.8 of all individuals heterozygous for a given SNP. To account for physical linkage, we restricted the final VCF file to one random SNP per locus.

We generated six SNP data sets (Table 1): (a) full taxon sampling (n = 82), (b) all taxa but downsampled to have a maximum sample size of four per population (n = 25), (c) as the former but including two S. trivirgatus subpopulation haplogroups (n = 29), (d) for the guttula haplogroup of Australian S. trivirgatus and southern New Guinean S. guttula (n = 24), (e) for only Australian S. trivirgatus (n = 44), and (f) for only New Guinean S. guttula (n = 20).

2.6 | Population structure and assignment

We used two approaches to determine basic population structure. First, we used SNMF (Frichot et al., 2014), implemented in the R package LEA 2.2.0 (Frichot & François, 2015), to infer the best-fit number of populations (k) and construct assignment plots with admixture coefficients. We checked for consistency of k under varying values of the regularization parameter alpha (1, 10, and 100), in which higher values penalize admixture more than lower values. Once we chose an optimal value of alpha based on a preliminary investigation of which value minimized the cross-entropy criterion, we performed 100 replicates for each value of k tested. Due to the unstable nature of estimating k (Kalinowski, 2011) and given uneven sampling, we include and interpret values of k greater than the one inferred by the cross-entropy criterion. Second, we used the gPca method within ADEGENET (Jombart, 2008) to conduct a principal component analysis on the data sets including only New Guinean S. guttula and Australian S. trivirgatus, respectively.

2.7 | Tests of gene flow

We generated two species trees with the SNP data. First, in SNAPP v1.4.2 (Bryant et al., 2012), we included up to four individuals per population, except in cases where less than four individuals were available (i.e., n = 3 of S. t. bimaculatus, and n = 2 for each of S. t. diadematus, S. t. nigrimentum, and S. t. trivirgatus). We ran two million generations and discarded the first quartile as burnin; we calculated mutation parameters automatically rather than inferring them from population size estimates. This tree enabled an alternate topology to better assess the validity of ABBA/BABA tests (described in detail below). Second, we used TREEMIX v1.13 (Pickrell & Pritchard, 2012), which uses the covariance of populations’ allele frequencies to estimate an ML species tree and then identifies species that are more closely related than can be explained by the relationships in the species tree. These species pairs are thus candidates for migration events (i.e., migration edges in the phylogeny) and we added migration edges until we observed an increase in likelihood of <1. First, we performed an unsupervised run of TREEMIX, where no prior migration edges were tested. To observe the effects of migration events

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>Numbers of samples included and SNPs recovered in each dataset used for population-level analyses. Note that although 4,051 total SNPs were used to create input for ABBA/BABA analysis, only 2,756 were informative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset</td>
<td>N samples</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Full sampling (with S. verticalis outgroup)</td>
<td>82</td>
</tr>
<tr>
<td>All major populations (n ≤ 4)</td>
<td>25</td>
</tr>
<tr>
<td>All major populations and a second Australian S. trivirgatus group (n ≤ 4)</td>
<td>29</td>
</tr>
<tr>
<td>One subpopulation each of S. guttula (southern New Guinea) and S. trivirgatus (S. guttula haplogroup sans S. t. albiventris)</td>
<td>24</td>
</tr>
<tr>
<td>All samples of Australian S. trivirgatus</td>
<td>44</td>
</tr>
<tr>
<td>All samples of New Guinea S. guttula</td>
<td>20</td>
</tr>
</tbody>
</table>
supported by ABBA/BABA tests and mitochondrial discordance, we also performed supervised runs of treemix by adding an a priori migration edges using the -cor_mig and -climb flags. Specifically, we manually tested four edges: (a) between New Guinean S. t. diadematus to Gag Island S. guttula; S. t. melanopterus to S. t. diadematus; southern New Guinea S. guttula to Australian S. trivirgatus; and S. t. melanopterus to S. guttula from the Louiisiades. This allowed for testing the hypothesis of introgression to explain phylogenetic uncertainty between analyses.

To test for evidence of modern or historic gene flow between S. guttula and Australian S. trivirgatus, we used ABBA/BABA tests (Durand et al., 2011). Specifically, we tested a priori migration edges of interest, as well as instances of non-sister introgression identified by our treemix analyses. The ABBA/BABA test generates a D statistic with a ladderized tree with four tips (tips labelled P1–4), where positive values imply that there is gene flow between P2 and P3, and negative values imply gene flow between P1 and P3. The magnitude of D is proportional to the amount of gene flow between P3 and the other population, but this observation can be confounded by gene flow between P1 and P2 (Durand et al., 2011). We focused on four major migration edges: (a) between New Guinean S. guttula and Australian S. trivirgatus, (b) between Louiisiades S. guttula and S. t. melanopterus, (c) between Gag Island S. guttula and S. t. diadematus, and (d) between S. t. melanopterus and S. t. diadematus. In addition to these focused tests, we tested additional ABBA/BABA patterns to demonstrate the local nature of this gene flow in cases where taxa had multiple subspecies or populations (see Table S2 for a full list). We generated input files using scripts by Simon Martin (https://github.com/simonhmartin/genomics_general) and ran the ABBA/BABA tests in r (R Development Core Team, 2013). We performed 1,000 bootstrap replicates to generate a distribution of values for both statistics, and we calculated p-values with the proportion of replicates less than or equal to 0.05.

2.8 | Demographic analyses

We performed demographic analyses and model testing with DAdi v2.0.3 (Gutenkunst et al., 2009) to test different models of gene flow and admixture between Australian S. trivirgatus and New Guinean S. guttula (visualized in Figure S1). Due to notable population structure within both taxa, we chose the largest and most relevant populations from each: S. guttula from southern New Guinea and S. trivirgatus (excluding S. t. albibrevis) within the S. guttula mitochondrial haplogroup. We generated input files for this analysis using EASYSFS (Oswald et al., 2017; https://github.com/isaacovercast/easySFS) to generate and downsample (for even sample size) a site frequency spectrum (SFS) for DAdi. Once the SFS was generated, we used it to test ten different models (visualized in Figure S1). The null model did not allow migration or admixture between populations after divergence. The next three models allowed for continuous, unidirectional (in either direction) or bidirectional migration. The final six models included a discrete unidirectional or bidirectional admixture event, as well as either including or not including continuous gene flow between populations. We performed 50 optimization runs with perturbed parameters, whose bounds were universally set across models to ensure that no model was limited by its parameters. We chose the highest likelihood run from a given model for comparison and compared these models using an Akaike information criterion (AIC) weight framework using likelihoods calculated in DAdi (Rougeux et al., 2019). Following Winker et al. (2018), we blasted our VelvetOptimiser reference assembly FASTA (see SNP calling from UCEs section above) against genomes of Melopsittacus undulatus and Acanthisitta chloris, and used the date of their most recent common ancestor (60.5 and 53 Ma, respectively; Claramunt & Cracraft, 2015) to estimate substitution rates for the sequences from which SNPs were called to interpret demographic parameters. The final effective sequence length and mutation rate were 834,259 bp (the length of the entire reference FASTA) and $5.6 \times 10^{-10}$ mutations per site per generation, respectively. A generation time of two years was used, as in Ficedula (Smeds et al. 2016), due to the uncertainty in actual generation time of our study system. Note that our primary parameter of interest, admixture proportion, is not affected by these parameters.

2.9 | Detecting positive selection in the mitogenome

We tested for positive selection in each of the 13 mitochondrial genes with several methods. Using the McDonald-Kreitman test (MK test; McDonald & Kreitman, 1991), we derived p-values from a Chi-squared analysis of a contingency table of fixed and polymorphic mutations. These mutations can either be synonymous or non-synonymous between the focal group and the outgroup, S. verticalis. We examined the three mitochondrial clades: S. trivirgatus with the S. guttula mtDNA, S. trivirgatus with the “true” trivirgatus mtDNA, and S. guttula from southern New Guinea. For each population, we computed the Chi-squared value and Neutrality Index (NI; Rand & Kann, 1996) with the standard MK test web portal (http://mkt.uab.es/mkt/; Egea et al., 2008). Taken together, these two statistics test whether any mitochondrial group showed evidence of gene-level positive selection that significantly deviated from neutrality. Furthermore, the proportion of fixed amino acid substitutions was calculated to see if the genes in one group were enriched for substitutions relative to others. In addition, we used CODEML (Yang, 2007), as implemented in EASYCODEML (Gao et al., 2019), to test for natural selection in each of the 13 individual mitochondrial coding genes and accompanying gene trees with the site-specific model. This model allows the ratio of nonsynonymous to synonymous substitution rates ($\omega$) to vary among sites (but not branches) to test for site-specific evolution in the context of a phylogeny. For each gene region, we evaluated which of six codon substitution models (M0, M1a, M2a, M3, M7, M8) was favored using likelihood ratio tests.
To address the adaptive capture hypothesis, we also performed three similar statistical tests to detect a selective sweep in the mtDNA. These three tests are Tajima's D (Tajima, 1989) and Fu and Li's F* and D* (Fu & Li, 1993). All three tests were performed in DNAsp 6 (Rozas et al., 2017). These tests should return a significantly negative value if a historical bottleneck or selective sweep has occurred. In order to correct for demography, we performed this test on all three mitochondrial clades tested in the prior section. If adaptive capture has occurred, we would expect a significantly negative value of captured S. guttula mtDNA that is also less than that of native S. trivirgatus and S. guttula mtDNA.

3 | RESULTS

3.1 | Sampling

We recovered an average of 4,186 UCEs (median 4,242, range: 3,685–4,441) per sample. The 75% complete matrix (i.e., 75% of taxa were present per UCE alignment) comprised 4,269 loci and 3,970,857 bp, of which 101,808 bp were variable and 54,002 bp were parsimony-informative characters. We recovered full mitogenomes from 79 individuals totalling 16,983 bp, but we reduced our final mitogenome data matrix for phylogenetic analysis to only the 13 coding genes (11,404 total bp). We removed two toepad-derived samples (S. t. nigrimentum AMNH654697 and AMNH654694) from all downstream UCE- and SNP-based analyses due to excessively long branch lengths and incorrect topologies that were apparent in preliminary RAxML analyses; however, we kept these samples in the ND2 haplotype network because all recovered mtDNA was suitable for analysis. The six SNP data sets that we produced for this study, including numbers of SNPs and their specific uses, are outlined in Table 1. Note that we produced data sets without singletons for sNMF analyses because their inclusion has been shown to decrease the effectiveness of model-based analyses of population structure (Linck & Battey, 2019).

3.2 | Phylogenetics

Mitochondrial and nuclear data showed conflicting relationships across the Symposiachrus trivirgatus and S. guttula species complexes. The mitochondrial tree (Figure 1) and haplotype network (Figure S2) indicated that there are two haplogroups within Australian S. trivirgatus: those with the S. guttula mitochondrion and those we infer to have the "true" S. trivirgatus mitochondrion (hereafter referred to as the "guttula St" and "trivirgatus St" haplogroups within S. trivirgatus, respectively; Figure 1b,c). Median uncorrected p-distances of mitochondrial ND2 and nuclear UCEs are reported in Table 2. ND2 genetic distances ranged from 0.96% to 5.8% and UCE genetic distances ranged from 0.18% to 0.34%. As expected, guttula St and trivirgatus St were 5.0% divergent in ND2, but only 0.18% divergent in the nuclear UCE data set. Overall, the mitogenome tree was largely similar to that of Andersen et al. (2015) in that S. trivirgatus was paraphyletic and comprised five clades (colours as in Figure 1b):

1. S. t. trivirgatus of the Lesser Sundas (red), which was sister to
2. Australian trivirgatus St haplogroup clade (orange).
3. Maluku S. t. nigrimentum, S. t. diadematus, and S. t. bimaculatus (light purple, dark purple, and pink, respectively).
4. S. t. melanopterus, endemic to offshore islands of far south-eastern Papua New Guinea (yellow), distinct from all other S. trivirgatus and sister to S. menckei, an endemic species from Mussau Island in the Bismarck Archipelago.
5. Remaining Australian S. trivirgatus samples (guttula St haplogroup; blue) as sister to S. guttula from New Guinea.

Within S. guttula, we found some phylogenetic structure, including a clade of two samples from the Louisiade Archipelago and three samples from East Sepik and Madang provinces in northwest Papua New Guinea, to the exclusion of all other S. guttula samples (green clades; Figure 1b); however, recall that we did not sample from western mainland New Guinea. Finally, S. guttula from Gag Island—offshore from New Guinea’s Vogelkop Peninsula (hereafter Gag guttula)—formed a clade that was not sister to other S. guttula. The ND2 haplotype network (Figure S2) mirrored the relationships in the mitogenome phylogeny.

Patterns of paraphyly in the mitogenome tree largely disappeared in concatenated and species trees inferred from nuclear data. Our RAxML tree, based on a partitioned, concatenated analysis of UCEs (Figure 1a) indicates that all S. trivirgatus (barring the Maluku subspecies) formed a clade. Samples belonging to the two mtDNA haplogroups of Australian S. trivirgatus (paraphyletic in the mtDNA tree) were indistinguishable in the UCE tree. Nominate S. t. trivirgatus was sister to all Australian individuals and S. t. melanopterus was the first branch in this clade, sister to the rest. We also recovered novel relationships in the nuclear tree with respect to Gag guttula, which was strongly supported as sister to the Maluku S. trivirgatus samples, rather than as an independent lineage subtended from a polytomy within the tree. And, unlike the mitogenome tree, all four far south-eastern Papua New Guinea S. guttula samples formed a clade to the exclusion of those from northwest Papua New Guinea.

Species trees, estimated with either full loci or SNPs, show similar results to the concatenated analysis: (a) both S. trivirgatus mitochondrial haplogroups are monophyletic, and (b) the Maluku subspecies are more closely related to S. guttula than nominate and Australian S. trivirgatus (Figure 2, S3). Yet, the only significant differences between species tree analyses was the equivocal topology regarding the sister lineage of the closest relative of the clade of Maluku S. trivirgatus subspecies. The SVDquartets analysis indicated with moderate support (72%–73% bootstrap values) that the clade of Maluku subspecies was closest to Gag guttula, which was in turn sister to the rest of S. guttula (Figure 2) and thus supported the RAxML topology. The most likely SNAPP species tree supported all of S. guttula (including Gag Island samples) as the sister clade to the Maluku subspecies (88.8% of trees in the posterior distribution);
however, two alternative topologies were represented in the posterior distribution—either Gag guttula or mainland S. guttula were the sister lineage to Maluku S. trivirgatus (5.6% or 5.3% of trees, respectively). Note that S. t. melanopterus was inferred as the outgroup lineage in SNAPP; however, this result may be unreliable due to our inability to include an actual outgroup given insufficient outgroup sample size. Finally, our treeMix results mostly supported the sister relationship of Gag guttula and Maluku S. trivirgatus populations

**TABLE 2** Median uncorrected p-distances between six groups of *Symposiachrus* monarch-flycatchers. Numbers above the diagonal are distances in the nuclear UCE data; numbers below the diagonal are based on the mitochondrial ND2 gene

<table>
<thead>
<tr>
<th>Mitochondrial group</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 S. t. trivirgatus</td>
<td></td>
<td>0.25%</td>
<td>0.29%</td>
<td>0.34%</td>
<td>0.26%</td>
<td>0.33%</td>
</tr>
<tr>
<td>2 trivirgatus St haplogroup</td>
<td>2.40%</td>
<td></td>
<td>0.28%</td>
<td>0.32%</td>
<td>0.18%</td>
<td>0.31%</td>
</tr>
<tr>
<td>3 S. t. bimaculatus, S. t. diadematus, S. t. nigriimentum</td>
<td>5.20%</td>
<td>5.80%</td>
<td></td>
<td>0.30%</td>
<td>0.28%</td>
<td>0.26%</td>
</tr>
<tr>
<td>4 S. t. melanopterus</td>
<td>4.70%</td>
<td>4.80%</td>
<td>4.30%</td>
<td></td>
<td>0.32%</td>
<td>0.34%</td>
</tr>
<tr>
<td>5 guttula St haplogroup</td>
<td>4.50%</td>
<td>5.00%</td>
<td>4.00%</td>
<td>3.70%</td>
<td></td>
<td>0.31%</td>
</tr>
<tr>
<td>6 S. guttula</td>
<td>4.37%</td>
<td>4.71%</td>
<td>4.04%</td>
<td>3.55%</td>
<td>0.96%</td>
<td></td>
</tr>
</tbody>
</table>

**FIGURE 2** Genomic exploration of population structure within the *Symposiachrus trivirgatus/guttula* complex. (a) Population assignment plots generated in sNMF for a range of k values from 5–8. (b) Principal components analysis of SNP data, showing the first two PC components (contribution of PC1 and two axes are 17.6% and 15.8%, respectively). A full table of the first four PC axes is reported in Table S3. (c) Species tree based on 4,269 UCE loci and estimated in SVDQuartets, all nodes are fully supported unless noted otherwise. (d) Species tree based on 1,855 SNPs and estimated in SNAPP
Figure 3  Population structure of Australian Symposiachrus trivirgatus. (a) Sampling map and ranges of the three S. trivirgatus subspecies in Australia, including two immature birds (possible migrants from Australia) captured in Papua New Guinea. Colours represent the three genetic clusters found in our analyses: teal = S. t. albiventris (all but two of which have the S. guttula mtDNA, exceptions are noted below); orange = individuals with “true” S. trivirgatus mtDNA; blue = nonalbiventris individuals with S. guttula mtDNA. The latter two genetic clusters are not delineated by subspecies, as gouldii and melanorrhous are morphologically indistinguishable, and they separate genetically in all analyses of population structure. (b) Population structure of S. trivirgatus based on sNMF analyses, including results from k = 2–3. When k = 2, the guttula haplogroup is determined to be distinct (i.e., blue individuals), but when k = 3, S. t. albiventris (teal) and the orange group separated. Subspecific identity of the three left-most samples indicated by a dashed line was unclear based on phenotype of specimens; however, they were assigned genetically to the S. t. albiventris population (from left to right: ANWC B57069, ANWC B57291, and ANWC B57103). (c) Principal components of Australian S. trivirgatus based on the SNP dataset. Six points are noted in this figure: a (ANWC B43114) and b (ANWC B43078) are identified as the only sampled albiventris individuals with the “correct”; trivirgatus mitochondrial; points a, c (ANWC B28832), d (ANWC B31374), e (ANWC B31273), and f (ANWC B43719) are hybrid individuals. See Figure S5 for a similar PCA with these outliers included in their a priori groups. Original artwork of S. t. albiventris by Madison E. Mayfield

3.3  Population structure and assignment

Our sNMF analyses revealed population structure across this study system and that all Australian S. trivirgatus are distinct in nuclear DNA from S. guttula, regardless of mitochondrial haplogroup (Figure 2). Because k optimization is often conservative and can be unreliable (Kalinowski, 2011), we included analyses from a range of k (5–8), which parsed ever more geographically circumscribed populations as k increased. For example, at k = 5, all populations except the Maluku, Lesser Sundas, and Australian S. trivirgatus were distinct; whereas at k = 8, all but two populations were assigned as distinct (Figure 2b). This analysis did not find clear population structure within Australian S. trivirgatus. However, the S. trivirgatus with the “true” trivirgatus mitochondrial DNA showed a mixed assignment, which suggested that more detailed analysis within Australian S. trivirgatus was needed.

When testing for recent admixture between S. guttula and guttula St, we inferred minor levels of admixture (all proportions ≤0.1). This was probably due to noise in the data, introduced by population structure within S. trivirgatus, not higher order backcrosses (Figure S4). Intraspecific population structure analyses between S. t. albiventris, trivirgatus St, and guttula St (Figure 3) confirms population structure in Australian birds. Our results indicated three genetic clusters with few admixed individuals: one with the guttula St haplogroup, one with the trivirgatus St haplogroup, and a separate cluster of S. t. albiventris from the Cape York Peninsula. This latter subspecies, which primarily groups together in both nuclear and mitochondrial trees and are mostly assigned to the guttula St haplogroup (Figure 1), is a genetically distinct population based on both sNMF and genetic PCAs (Table S3, Figure 3, S5). The two other genetic groups are not obviously delineated by Australian S. trivirgatus subspecies, S. t. gouldii and S. t. melanorrhous, but rather by mitochondrial haplogroup (Fisher’s exact test, p < .001). The trivirgatus St haplogroup cluster was more similar to S. t. albiventris (overlapping 95% inertia ellipses; Figure 3c), for which all but two S. t. albiven-
tris were in the guttula St haplogroup clade (Figure 1b); the most distinctive grouping was the guttula St haplogroup nuclear cluster (Figure 3c). Despite their distinctiveness, there is still ongoing gene flow between all three genetic clusters, as shown by three putative F1 or F2 hybrids (see labeled individuals in Figure 3), which indicates that these populations are not reproductively isolated. Both the S. t. albiventris and guttula St haplogroup clusters differed significantly from an expectation of random mtDNA haplogroup distributions (Chi-squared test, p < .01), but did not differ from each other (Fisher’s exact test, p = 1). Although the 95% inertia ellipse of the trivirgatus St haplogroup cluster did not separate entirely from the S. t. albiventris
group in the PCA, their distinctness was supported by sNMF and significantly different (Fisher’s exact test, $p < .001$) distributions of mitochondrial haplogroups between the two clusters in PCA space (Figure 3c, S5). Superficially, the mitochondrial haplogroups do not correspond well to our sampling map (Figure 1). However, when only breeding-season specimens are considered, both *Tririgatus* St haplogroup and *guttula* St haplogroup largely correspond to the breeding ranges of *melanorrhous* and *gouldii*, respectively (Figure S6).

Despite no phenotypic variation or subspecies being recognized in *guttula* (Beehler & Pratt, 2016), we observed population structure across its populations in eastern mainland New Guinea and offshore islands. Conversely, in both mitochondrial and nuclear trees, *Gag guttula* consistently grouped on its own and was separate from mainland New Guinea samples (Figures 1–2, S7). Though $k = 1$ was our most supported assignment, further exploration of New Guinean *guttula* uncovered five geographically relevant subpopulations (Figure S7). We first observed distinct population structure in samples from the Louisiade Archipelago, which are sympatric with *S. t. melanopterus*, then with Normanby Island in the D’Entrecasteaux Archipelago (yet this population was not distinct in the PCA; Figure S7). At $k = 4$, we found a divide north and south of the New Guinea Central Cordillera, and weaker population structure on either side of the Huon Peninsula at $k = 5$. Lacking samples from the Huon Peninsula, we cannot determine where this break lies.

### Tests of gene flow and demographic analyses

Despite evidence of a mitochondrial capture event, we did not detect significant modern gene flow between the Australian *guttula* St haplogroup or mainland New Guinea *S. guttula* (Figure 4a). This negative result in the face of evidence for mitochondrial capture can probably be explained by the reduced power of ABBA/BABA tests in cases with low levels of admixture and reduced representation of the genome as in this study (Durand et al., 2011). Instead, we revealed a complicated history of gene flow elsewhere within this species complex. For example, ABBA/BABA detected significant gene flow between *S. t. diadematus* and Gag *guttula* (Figure 4b) and also in the Louisiade Archipelago, where *S. t. melanopterus* and a population of *S. guttula* occur sympatrically (Figure 4c); however, gene flow was not detected between *S. t. melanopterus* and other *S. guttula* (Table S2). Unexpectedly, we also detected significant gene flow between *S. t. melanopterus* and *S. t. diadematus* (Figure 4d).

**Figure 4** ABBA/BABA tests shown by topologies and visualized using density plots of values of D obtained by bootstrapping. A value greater than zero is indicative of either gene flow or an incorrect topology tested. Topologies that are supported by our RAxML and/or SNAPP trees are denoted above the trees and $p$ and D values are listed below. Note that certain tests, denoted by the green topologies (right) are supported by one method of phylogenetic analysis, but the orange topology (left) was supported by all phylogenetic methods used in this study. Therefore, significant results from those topologies are strong evidence for gene flow, as incorrect topologies can result in falsely rejecting the null hypothesis of no gene flow. (a) The test for gene flow between Australian *S. tririgatus* (the *guttula* St haplogroup) and New Guinea *S. guttula* (southern population), demonstrating no statistically significant nuclear introgression despite mitochondrial capture. (b) The test for gene flow between *S. t. diadematus* and Gag Island *guttula*, showing significance for both reciprocal tests. (c) The test for gene flow between *S. t. melanopterus* and *S. guttula* from the Louisiade Archipelago, where the two species co-occur. (d) Test for gene flow between *S. t. melanopterus* and *S. t. diadematus*

Our *treemix* analysis did not support gene flow between mainland New Guinean *guttula* and the Australian *guttula* St haplogroup. Though not supporting gene flow between the expected taxa, *treemix* did indicate historical gene flow in this system elsewhere. When unsupervised, *treemix* produced a topology concordant with RAxML (Figure S3A) and inferred gene flow between *S. t. melanopterus* and the ancestor of the Maluku *S. tririgatus* populations (Figure S3B; not significant by the jackknifing algorithm, $p = .18$). Our supervised analysis supported gene flow from *S. t. diadematus* to Gag *guttula* ($p = 0.03$; admixture proportion $= 0.217$), but resulted in a major topological rearrangement (Figure S3c). Other supervised tests of historical gene flow were not statistically significant and resulted in no topological change, which highlights the authenticity of the inferred gene flow from *S. t. diadematus* to Gag *guttula*, as opposed to an artifact due to manually adding a migration edge.

However, our demographic analyses supported at least one recent, discrete admixture event between *guttula* and *St* St, the mitochondrially introgressed individuals of *S. tririgatus*. *Dadi* indicated the best fit model was one with recent, bidirectional admixture without ongoing migration (Table 3). All well-supported models (AIcw of admixture models $= .991$) included a discrete gene flow event; however, we acknowledge that it was difficult to distinguish between different models that include discrete gene flow events. For models with discrete and continuous gene flow in the same direction, the amount of inferred gene flow in that direction was nearly nonexistent (less than 1 migrant every 100 generations). The *S. guttula* population size was approximately 10x larger than the included *S. tririgatus* population (Table 3), and *Dadi* indicated low admixture proportions for each direction of admixture ($\sim$1.0–2.5% for each model) that occurred approximately 5,000 years ago. However, estimates of divergence time and population size in *Dadi* for regions with high linked selection can perform poorly, whereas estimates of gene flow tend to perform well (Ewing & Jensen, 2016).

### Selection in the mitogenome

All tests for selection in the mitogenome did not indicate positive selection or a sweep for *S. tririgatus* individuals with *S. guttula* mtDNA (Figure S8). Specifically, the McDonald-Kreitman (MK) tests and associated calculation of the neutrality indices for all mitochondrial genes found no evidence for positive selection (Table S4). All significant values of the MK test had a neutrality index greater than
TABLE 3 Results of Dadi analyses, including estimates of parameters for population sizes, divergence times, migration rates (in number of migrants), admixture proportion, and admixture timing

<table>
<thead>
<tr>
<th>Model</th>
<th>S. guttula population</th>
<th>S. trivirgatus population</th>
<th>Divergence time (years)</th>
<th>Time of admixture (years)</th>
<th>Admixture proportion of S. guttula into S. trivirgatus genome</th>
<th>Number of S. trivirgatus migrants into S. guttula per generation</th>
<th>Number of S. guttula migrants into S. trivirgatus per generation</th>
<th>AIC</th>
<th>AICw</th>
</tr>
</thead>
<tbody>
<tr>
<td>No migration</td>
<td>1,418,680</td>
<td>117,277</td>
<td>490,141</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>508.0</td>
<td>1.0E-06</td>
</tr>
<tr>
<td>Continuous gene flow from S. trivirgatus</td>
<td>1,307,927</td>
<td>114,211</td>
<td>520,005</td>
<td>-</td>
<td>-</td>
<td>4.7E-01</td>
<td>-</td>
<td>491.1</td>
<td>4.7E-03</td>
</tr>
<tr>
<td>Continuous gene flow from S. guttula to S.</td>
<td>1,341,420</td>
<td>106,649</td>
<td>511,957</td>
<td>-</td>
<td>-</td>
<td>6.3E-02</td>
<td>-</td>
<td>498.2</td>
<td>1.4E-04</td>
</tr>
<tr>
<td>Continuous bidirectional gene flow between</td>
<td>1,294,535</td>
<td>109,161</td>
<td>532,143</td>
<td>-</td>
<td>-</td>
<td>3.6E-01</td>
<td>3.0E-02</td>
<td>491.6</td>
<td>3.7E-03</td>
</tr>
<tr>
<td>Single admixture event from S. guttula to S.</td>
<td>1,383,458</td>
<td>101,336</td>
<td>523,241</td>
<td>3.046</td>
<td>2.5E-02</td>
<td>-</td>
<td>3.4E-01</td>
<td>483.8</td>
<td>1.8E-01</td>
</tr>
<tr>
<td>Single admixture event from S. trivirgatus</td>
<td>1,327,902</td>
<td>114,172</td>
<td>503,848</td>
<td>19,361</td>
<td>-</td>
<td>2.0E-02</td>
<td>-</td>
<td>490.4</td>
<td>6.8E-03</td>
</tr>
<tr>
<td>Single admixture event from S. trivirgatus</td>
<td>1,314,031</td>
<td>110,430</td>
<td>486,650</td>
<td>42,425</td>
<td>-</td>
<td>2.0E-02</td>
<td>3.2E-03</td>
<td>493.4</td>
<td>1.5E-03</td>
</tr>
<tr>
<td>Single bidirectional admixture event</td>
<td>1,405,950</td>
<td>101,890</td>
<td>515,915</td>
<td>4,957</td>
<td>2.5E-02</td>
<td>1.3E-02</td>
<td>-</td>
<td>481.0</td>
<td>7.3E-01</td>
</tr>
<tr>
<td>Single bidirectional admixture event, cont</td>
<td>1,401,649</td>
<td>101,280</td>
<td>520,412</td>
<td>4,627</td>
<td>2.5E-02</td>
<td>1.0E-03</td>
<td>2.9E-01</td>
<td>486.0</td>
<td>6.0E-02</td>
</tr>
</tbody>
</table>

Our genomic investigation of the Spectacled Monarch Symposiachrus trivirgatus and Spot-winged Monarch Symposiachrus guttula complex in Australia, Papua New Guinea, and Indonesia revealed a complex evolutionary history. This includes a clear instance of massively discordant mitochondrial introgression (MDMI, or mitochondrial capture) from S. guttula into its non-sister species (Figure 1). No evidence of a selective sweep or selection for mitochondrial capture of S. guttula by its non-sister species (S. trivirgatus) was found. S. guttula shows strong support for mitochondrial capture of S. guttula by its non-sister species (S. trivirgatus) (Table S4, Figure S8). Nuclear DNA also showed statistically significant results for the M7 versus M8 model comparison, but not M8a versus M8, which yields fewer false positives (Gao et al., 2019; Swanson et al., 2003; Wong et al., 2004), indicating significant positive selection in both the M7 versus M8 and M8a versus M8 model comparisons and each contained one positively selected site (Table S5). Significant positive selection (p < 0.05) at up to three other loci and significantly negative (p < 0.05) at five loci (Table S5) were both lower and less significant than those of M8a versus M8 model comparisons. COI showed statistical support (p < 0.05) for a selective sweep in the mitochondria of S. guttula at opposite MTN4 (Position 41, N to S; ND2: Position 31, T to A) with a posterior probability over 0.95 (Position 472, V to I). However, there was no phylogenetic signal to the capture of the S. guttula genome, and our analysis did not support a selective sweep or selection for mitochondrial introgression (MDMI, or mitochondrial capture). Furthermore, the progression of far southeastern Papua New Guinea and Indonesia revealed a complex evolutionary history. This includes a clear instance of massively discordant mitochondrial introgression (MDMI, or mitochondrial capture) from S. guttula into its non-sister species (S. trivirgatus) (Figure 1). No evidence of a selective sweep or selection for mitochondrial capture of S. guttula by its non-sister species (S. trivirgatus) was found. S. guttula shows strong support for mitochondrial capture of S. guttula by its non-sister species (S. trivirgatus) (Table S4, Figure S8). Nuclear DNA also showed statistically significant results for the M7 versus M8 model comparison, but not M8a versus M8, which yields fewer false positives (Gao et al., 2019; Swanson et al., 2003; Wong et al., 2004), indicating significant positive selection in both the M7 versus M8 and M8a versus M8 model comparisons and each contained one positively selected site (Table S5). Significant positive selection (p < 0.05) at up to three other loci and significantly negative (p < 0.05) at five loci (Table S5) were both lower and less significant than those of M8a versus M8 model comparisons. COI showed statistical support (p < 0.05) for a selective sweep in the mitochondria of S. guttula at opposite MTN4 (Position 41, N to S; ND2: Position 31, T to A) with a posterior probability over 0.95 (Position 472, V to I). However, there was no phylogenetic signal to the capture of the S. guttula genome, and our analysis did not support a selective sweep or selection for mitochondrial introgression (MDMI, or mitochondrial capture). Additionally, CodeML analyses showed little support for a selective sweep in the mitochondria of S. guttula (Table S5).
the details of our findings in terms of the four hypotheses proposed at the outset: cryptic species, gene flow, neutral capture, and adaptive capture. We first discuss interactions between Australian S. trivirgatus and S. guttula and then between Indonesian S. trivirgatus and S. guttula.

4.1 | Australian S. trivirgatus and S. guttula: multiple hypotheses supported?

Symposiachrus trivirgatus, sampled from across Australia, was assigned to either of two highly divergent nonsister haplogroups across the Symposiachrus mitogenome tree (Figure 1). UCE-based analyses of nuclear DNA found no similar patterns, S. trivirgatus and S. guttula being reciprocally monophyletic with respect to each other (with the exception of the Maluku taxa, which are a separate monophyletic clade). There is no signal of nuclear introgression between S. guttula and S. trivirgatus that possess the S. guttula mitochondria (Figure 2, S4). They are affirmed as nonsister species and we can reject the cryptic species hypothesis in Australian S. trivirgatus.

To address the neutral hypothesis, we first consider our results on the timing of the capture. Our results are consistent with Pleistocene glacial cycles that facilitated geographic and genetic contact between S. guttula and S. trivirgatus. During these cycles, lower sea levels exposed the vast Sahul Shelf connecting northern Australia and New Guinea up to eight times (Lambeck et al., 2002; Lambeck & Chappell, 2001; Voris, 2000) and for much of the last 100,000 years (review in Joseph et al., 2019; Voris, 2000). By promoting connectivity between faunas of Australia and New Guinea (Heinsohn & Hope, 2006; Peñalba et al., 2019), repeated episodes of range expansion, secondary contact, and gene flow between S. guttula and S. trivirgatus would have been possible. Based on ND2 substitution rates derived from birds (2.1%–3.3%; Lerner et al., 2011; Weir & Schluter, 2007), our results suggest gene flow could have occurred between Australian S. trivirgatus and New Guinean S. guttula as recently as 5,000 years ago (Table 3) or as long ago as up to 200,000–400,000 years ago. The uncertainty is attributed to repeated bouts of exposure of the Sahul Shelf as sea levels rose and fell. Nonetheless, our most recent inferred estimate for an admixture event is consistent with simulation studies suggesting that genomic data will reflect only the most recent glacial cycles (Linck & Battey, 2019). The best model supported bidirectional historic admixture but the strongest supported directionality from S. guttula into S. trivirgatus. This again is consistent with S. trivirgatus having expanded into the range of S. guttula. On balance, a neutral capture implies a novel corollary: extinction of S. guttula from an unknown proportion of its range on the Australian mainland or the now submerged Sahul Shelf. To the extent that we cannot reject neutrality, this capture is consistent with palaeoenvironmental data allowing for the expansion of the range of S. trivirgatus into that of S. guttula (Currat et al., 2008; Rheindt & Edwards, 2011). This could have resulted in a high frequency of S. guttula mitochondrial DNA in Australian S. trivirgatus populations and its near fixation in two genetic clusters. One of these clusters comprises all samples of S. t. albiventeris in northernmost Australia and the other is widespread across the eastern coast of Australia (Figures 1 and 3).

Next, we note that our tests for selection found little to no evidence for positive selection acting on the mitochondria since the capture event for the guttula St haplogroup. Superficially, this is consistent with a neutral capture. Again, however, we note the difficulties posed by Bonnet et al. (2017) in rejecting adaptive mechanisms completely. The near fixation in two different nuclear clusters suggests the need to still consider adaptive mechanisms (Figure 3). The few mitochondrial genes with statistically significant deviations from neutrality were consistent with weakly deleterious mutations (Tables S3; Charlesworth & Eyre-Walker, 2008). This pattern is consistent with little to no positive selection acting on entire genes since the mitochondrial capture event and thus does not support an adaptive explanation.

Next we consider our tests for gene flow. The low proportion of admixture inferred by our demographic modeling may suggest a low likelihood of the neutral capture hypothesis. However, multiple instances of secondary contact caused by glacial cycles, the feasibility of which we have noted, can explain this pattern. If the guttula St mitochondrial haplotype introgressed into Australian populations of S. trivirgatus at a frequency of 2.5% (Table 3) each of the eight times that the Sahul Shelf was exposed during our timing estimates (e.g., between 10,000 and 400,000 years ago), we would expect an 18% chance of fixation by purely neutral demographic processes. Drift alone could therefore explain this observed pattern without a selective sweep. As discussed earlier, the directionality of introgression could further promote a neutral capture. Further, fixation in certain genetic clusters could be more likely if effective population sizes were low. This would decrease the expected time to fixation of an allele. For a given gene introgressed into a population, the expectation is that one allele (or, in this case, mitochondrial haplotype) will fix after an average of \( N_e \) generations. This process is accelerated for mtDNA, whose effective population size (\( N_e \)) is smaller and fixation time is fourfold faster than that of the nuclear genome.

To summarize, our extra sampling has confirmed a case of massively discordant mitochondrial introgression from New Guinean S. guttula into Australian S. trivirgatus. Due to their reciprocal monophyly as shown by nuclear data, we can reject the cryptic species hypothesis. We find, however, that while there is support for a neutral capture model and some gene flow, we cannot say with certainty that selective capture is not also at play in this system.

4.2 | Complex patterns of gene flow

Here we discuss three observations concerning populations on satellite islands at opposite ends of New Guinea: S. guttula on Gag Island immediately west of New Guinea, S. trivirgatus diadematus of the Maluku Islands further west of New Guinea, and S. trivirgatus...
melanopterus of the Louisiade Archipelago off New Guinea’s south-eastern tip. We plan a separate, more detailed discussion of species limits and biogeography of this group elsewhere (McCullough et al. unpublished data).

First, concatenated UCE data supported Gag Island S. guttula as sister to the monophyletic Maluku S. trivirgatus subspecies (bimaculatus, diadematus, and nigrimentum) and this group is in turn sister to mainland New Guinean S. guttula. Thus, both mtDNA (Andersen et al., 2015) and UCEs suggested Gag and mainland New Guinean S. guttula are paraphyletic. Species-tree and network analyses, however, were equivocal in determining the relationships of Gag Island S. guttula. ABBA/BABA tests detected significant gene flow between Maluku S. t. diadematus and Gag S. guttula (Figure 4b). Simultaneous analysis of phylogenetic relationships and introgression supported the equivocal relationships of Gag S. guttula was due to high levels of gene flow from S. t. diadematus or a related taxon (Figure S3C). These lines of evidence suggest that Gag S. guttula is a hybrid population, but this hypothesis lacks evidence from plumage characters.

Second, S. t. melanopterus of the Louisiades at the opposite end of mainland New Guinea is phenotypically and genetically distinct but comes into contact with S. guttula (Figure S6). We inferred some gene flow between the two in sympathy (Figure 4c). This is consistent with other cases of gene flow within this complex. Whether such gene flow is ongoing requires further study. However, we note that the subspecies-defining features of S. t. melanopterus (white behind auriculars and dark wing coverts; Figure 1) are similar to features of S. guttula, which may suggest a hybrid origin of this taxon.

Third, we detected significant and unexpected gene flow between S. t. melanopterus and S. t. diadematus, which are currently isolated from each other by the entirety of mainland New Guinea (Figure 4d). This could reflect Pleistocene connections across the exposed Sahul Shelf as modelled in similarly distributed species by Peñaiba et al. (2019). Alternatively, it may reflect a now extinct “ghost” taxon (Benham & Cheviron, 2019; Gopalakrishnan et al., 2018; Zhang et al., 2019) in the Louisiades. Such a ghost taxon would have been related to S. t. diadematus but would have to have been replaced by what is now S. t. melanopterus.

4.3 Future directions and conclusions

Monarch-flycatchers are an iconic, species-rich radiation that span the Indo-Pacific. They have played a major role in the development of speciation, biogeography, and community assembly theory (Mayr & Diamond, 2001). Here, we studied one species complex within one genus, Symposiachrus. We find that hybridization between S. guttula and S. trivirgatus, two non-sister species, has itself had a complex history of different outcomes concerning gene flow and mitochondrial capture. Although we found no evidence for adaptive introgression, and the captured mitochondrial haplogroup has not yet gone to fixation in all Australian S. trivirgatus, theoretical findings suggest that an adaptive capture is difficult to rule out (Bonnet et al., 2017). Further study of this and other mitochondrial discordance systems needs to consider joint modelling of nuclear and mitochondrial genomes and their interactions. This stresses that it may be inappropriate to consider a neutral model as the null to be rejected and that, instead, joint modelling and testing of neutral and adaptive scenarios will be more fruitful. One area of future study has been raised by Rougemont et al. (2020) and Kim et al. (2018); that is, how past introgression may maintain maladaptive alleles in such modelling is an exciting dimension to future work and should be possible as genomic resources for these monarch-flycatchers increase. Additionally, higher resolution genomic data will allow for more accurate testing of possible introgressed nuclear loci associated with captured mitochondrial haplotypes. Finally, at a broader scale, we suggest that denser geographic and genomic sampling of all Symposiachrus taxa from Wallacea to Melanesia will offer excellent opportunities to explore new frontiers and build on the rich history of evolutionary theory that is central to this region of the world, and could add to the story of widespread admixture found here.

ACKNOWLEDGEMENTS

We thank the following museums and collections staff for providing tissue and toe pad loans and other support: Paul Sweet, Tom Trombone, and Peter Capainolo, American Museum of Natural History; Robert Palmer and Alex Drew, Australian National Wildlife Collection, CSIRO, Canberra; Moe Flannery and Jack Dumbacher, California Academy of Sciences; Bulisa Iova, Papua New Guinea National Museum and Art Gallery; Mark Robbins, University of Kansas Biodiversity Institute; Ron Johnstone and Rebecca Bray, Western Australia Museum; Kristof Zyskowski, Yale Peabody Museum, and Jan Bolding Kristensen, Natural History Museum of Denmark. We thank the Ministry of Research and Technology/ National Agency for Research and Innovation (RISTEK- BRIN; Research permit SIP No: 026/SIP/FRP/1/2012 to K.A.J); the Ministry of Environment and Forestry, Republic of Indonesia; the Research Center for Biology, Indonesian Institute of Sciences (RCB-LIPI) for providing permits to carry out fieldwork in Indonesia and to export select samples. In addition, we thank Fangluan Gao for assistance with Codeml analyses. LJ acknowledges numerous scientific research permits granted to the Australian National Wildlife Collection over several decades and that supported this work in Australia and PNG. We thank the UNM Center for Advanced Research Computing (CARC) and Cyberinfrastructure for Phylogenetic Research (CIPRES), supported in part by the National Science Foundation, for providing high-performance computing resources used in this study. JMM acknowledges the UNM Grove Summer Research Scholarship. EFG acknowledges NSF’s Graduate Research Fellowship (DGE-650114). KAJ acknowledges a National Geographic Research and Exploration Grant (8853-10), the Dybron Hoffs Foundation, and the Corrit Foundation for financial support for fieldwork in Indonesia. KAJ is also grateful for the financial support received from the Villum Foundation (Young Investigator Programme, project no. 15560). This work was supported by the University of New Mexico and a National Science Foundation award to MJA [DEB-1557051]. We gratefully acknowledge the comments of four reviewers and helpful discussions with Timothée Bonnet.
AUTHOR CONTRIBUTIONS

DATA AVAILABILITY STATEMENT
Population genetic and phylogenetic input files will be uploaded to Dryad (https://doi.org/10.5061/dryad.hx3f8bgd7). Raw Illumina sequencing reads are available from the NCBI SRA (BioProject PRJNA698220). Mitogenomes are accessioned in NCBI GenBank (Accession #s MW623660–MW623735). All custom scripts generated for this study will be uploaded to Dryad, with updated versions available on github: https://github.com/ethangyllenhaal/assorted_scripts.

ORCID
Michael J. Andersen https://orcid.org/0000-0002-7220-5588
Jenna M. McCullough https://orcid.org/0000-0002-7664-3868
Ethan F. Gyllenhaal https://orcid.org/0000-0002-0835-8520
Xena M. Mapel https://orcid.org/0000-0002-7501-578X
Knud A. Janss https://orcid.org/0000-0002-1875-9504
Leo Joseph https://orcid.org/0000-0001-7564-1978

REFERENCES